Technical Report TR-67-46 May 1967

ON A UNIFIED CONVERGENCE THEORY FOR A CLASS OF ITERATIVE PROCESSES

by

Werner C. Rheinboldt
Research Professor

The work reported here was performed at the Computer Science Center of the University of Maryland and was in part supported under grant NsG 398 of the National Aeronautics and Space Administration and grant DA-AROD-31-124-G-676 of the U.S.Army Research Office - Durham,N.C.

Abstract

A general convergence theory is developed for iterative processes of the form $x_{k+1} = Gx_k$, $k=0,1,\ldots$; it is founded on certain nonlinear estimates for the iteration function G as well as on a socalled concept of majorizing sequences. This new approach reduces the study of the iterative process to that of a second order nonlinear difference equation.

The theory contains as special cases both the wellknown contraction theorem as well as the Newton-Kantorovich theorem, and, moreover, it encompasses all similar theorems given so far for approximate Newton processes of the form

$$x_{k+1} = x_k - B(x_k)Fx_k \ , \ k=0,1,..$$

Various generalizations of the theory applicable to non-stationary processes of the form $x_{k+1} = G_k x_k$, $k=0,1,..$, are also discussed.

A UNIFIED CONVERGENCE THEORY FOR A CLASS OF ITERATIVE PROCESSES

Werner C. Rheinboldt [1)]

1. Introduction

One of the central numerical techniques for solving a nonlinear operator equation $Fx = 0$ is Newton's method

$$(1.1) \qquad x_{k+1} = x_k - (F'(x_k))^{-1}Fx_k \quad , k=0,1,\dots \quad .$$

Here $F'(x)$ denotes the Frechet derivative of F at x. Partly to overcome possible numerical difficulties connected with the evaluation of F' at each step, and partly to simplify the solution of the linear problem associated with (1.1) for each k , approximate Newton processes of the general form

$$(1.2) \qquad x_{k+1} = x_k - B(x_k)Fx_k \quad , k=0,1,\dots$$

have received increasing attention. Here $B(x)$ is for each x a linear operator which is usually derived from or related to $F'(x)$. For example, the iterations studied by Ben-Israel [2],[3], Bryan [4], Lieberstein [11], and Zincenko [21], [22], are of this form and so are the socalled Newton-Gauss-Seidel methods considered by Ortega and Rheinboldt [15]. A generalization of (1.2) are the processes where instead of $B(x_k)$ only some sequence of linear operators B_k is given, i.e.,

$$(1.3) \qquad x_{k+1} = x_k - B_k Fx_k \quad , k=0,1,\dots \quad .$$

Some special results about iterations of this last type have been given by Bartle [1]. However, the form (1.3) encompasses almost all useful iterations and hence meaningful general results can probably only be expected once the B_k are specified more precisely.

When only metric properties of the underlying space are used, there are three broad classes of convergence theorems for methods such as those mentioned above. The point-of-attraction theorems begin with the existence of a solution x^* of $Fx = 0$ and assure convergence of the iterates x_k to x^* if only x_o is chosen sufficiently close to x^*. The most ideal and at the same time rarest theorems are the global ones where a 'large' domain D can be specified

---

[1)] Institute for Fluid Dynamics and Applied Mathematics, and Computer Science Center, University of Maryland, College Park, Maryland.

This work was in part supported by the U.S.National Aeronautics and Space Administration under grant NsG-398 and by the U.S.Army Research Office under grant DA-ARO-(D)-31-124 G 563.

and for any $x_o \in D$ the iterates converge to some solution of $Fx = 0$. Between these two extremes are those convergence theorems which begin with conditions about the initial approximation x_o and conclude from this that starting from that x_o the iterates converge to a solution. The best known prototypes for this last class are the contraction theorem and the Kantorovich proof of Newton's method (see e.g., [7], [8]).

In this paper a unified theory is developed for convergence results of the above third class applicable to a broad class of methods. Necessarily, of course, this theory contains as special cases the contraction- and the Newton-Kantorovich theorem, thereby refuting the observation frequently made that these two results are conceptually very different from each other. But more importantly, the theory also encompasses, as far as can be determined, all the theorems of the mentioned type given so far for processes of the form (1.2) and to some extent of the form (1.3).

The theory is founded on nonlinear estimates for the iteration function and on a socalled concept of majorizing sequences. This latter concept is based on a simple principle observed to underlie Kantorovich's majorant proof of Newton's method, while the nonlinear estimates occur in a natural way in the study of the processes (1.2) and represent a needed generalization of estimates used by Collatz [6] and Schröder [17], [18].

In its simplest form, the general convergence theory reduces the study of the iterative processes to that of a second order difference equation. In analysing the different known results about various processes, it turns out that the assumptions made always happen to lead to a simple solvable difference equation. Upon investigation of natural extensions of these simple difference equations, it is observed that the unified approach used here permits in several cases the generalization of these individual results - at the same time providing an insight into the various possibilities of proving many other similar results. Simultaneously, this approach makes it evident that a much deeper study of the resulting difference equations is now needed in order to arrive at more general and at the same time more widely usable convergence theorems for the processes (1.2) and (1.3).

The general convergence theory is presented in Section 2; then Section 3 gives as a first application a theorem for the process $x_{k+1} = Gx_k$, $k=0,1,\dots$, which in turn provides the basis for subsequent uniqueness results. Section 4 treats iterations of the form (1.2) and includes as special cases various theorems about Newton's method; finally, Section 5 develops an extension of the theory applicable to the general case (1.3).

I would like to express my thanks to Professor James M. Ortega of the University of Maryland for numerous enlightening discussions which helped crystallize several of the concepts developed here. In particular, Professor Ortega also found independently a short proof of Newton's method along the lines used here (see [16]).

2. Majorizing Sequences and Generalized Contractions

In [8] Kantorovich introduced a proof for the convergence of Newton's method (1.1) based on the socalled concept of a majorizing operator. In brief, an operator equation $x = Gx$ on a Banach space X is said to be majorized by the real equation $t = \varphi(t)$ if $\|Gx_o - x_o\| \leq \varphi(t_o) - t_o$ and $\|G'(x)\| \leq \varphi'(t)$ whenever $\|x - x_o\| \leq t - t_o$. Using this assumption, the convergence of the iterative process $x_{k+1} = Gx_k$ in X is deduced from that of the iteration $t_{k+1} = \varphi(t_k)$ on the real line. Although this approach proves to be effective for the study of Newton's method itself, it rests essentially on the requirement that the majorizing process have the same form as the underlying process, and this in turn is a severe limitation when it comes to the study of the general processes (1.2).

Actually, a closer study of this Kantorovich approach reveals that underlying it is a very simple principle. In order to describe it we introduce the following concept.

2.1 - Definition: Let $\{x_k\}$ be a sequence in the metric space X. Then a real non-negative sequence $\{t_k\}$ is said to majorize $\{x_k\}$ if

$$\rho(x_{k+1}, x_k) \leq t_{k+1} - t_k \quad , k=0,1,\ldots \quad .$$

Note that any majorizing sequence $\{t_k\}$ of $\{x_k\}$ is necessarily non-decreasing. If $\{t_k\}$ majorizes $\{x_k\} \subset X$, then for $m > k \geq 0$

$$(2.1) \qquad \rho(x_m, x_k) \leq \sum_{j=k}^{m-1} \rho(x_{j+1}, x_j) \leq \sum_{j=k}^{m-1} (t_{j+1} - t_j) = t_m - t_k \ .$$

Hence, if $\lim t_k = t^* < +\infty$ exists, then $\{x_k\}$ is a Cauchy sequence in X, and, therefore, if X is complete, $\lim x_k = x^*$ also exists and for $m \to \infty$ we obtain from (2.1) the error estimate

$$(2.2) \qquad \rho(x^*, x_k) \leq t^* - t_k \quad , k=0,1,\ldots \quad .$$

This is the above mentioned principle behind the Kantorovich proof of Newton's method. The idea of the majorizing sequence lies in the simple observation that if $\rho(x_{k+1}, x_k) \leq s_{k+1}$, then $\sum_1^\infty s_k < +\infty$ is a sufficient condition for the convergence of the sequence $\{x_k\}$ in the complete space X. The partial sums $t_o = 0$, $t_k = \sum_1^k s_j$, $k=1,2,\ldots$.

form a majorizing sequence of $\{x_k\}$. Since $\sum_0^\infty \rho(x_{k+1}, x_k)$ may be viewed as an upper bound on the 'total path length' of the sequence $\{x_k\}$, the majorizing principle says that when the total path length is bounded the sequence is convergent.

For this principle to be useful a mechanism is needed to obtain a majorizing sequence $\{t_k\}$ for given $\{x_k\}$. This in turn requires appropriate assumptions, either about the generating mechanism of the sequence $\{x_k\}$ or at least about the relation between succeeding members x_k and x_{k+1}. It turns out that in many cases when the x_k are defined by the process

$$x_{k+1} = Gx_k \quad , k=0,1,... \tag{2.3}$$

majorizing sequences can be constructed by solving a difference equation of the form

$$t_{k+1} - t_k = \varphi(t_k - t_{k-1}, t_k, t_{k-1}) \tag{2.4}$$

for given t_o and t_1.

To simplify the notation, the following class of functions shall be used.

<u>2.2</u> - <u>Definition</u>: A function $\varphi: Q \subset R^p \to R^1$ is said to be of class $\Gamma^p(Q)$ if it has the following properties: (a) The domain Q is a hypercube $Q = J_1 \times J_2 \times \cdots \times J_p$ where each J_i is an interval on $[0,\infty)$ containing 0, i.e., an interval of the form $[0,a]$, $[0,a)$, or $[0,\infty)$; (b) φ is non-negative and isotone on Q, i.e., if $(u_1^{(i)}, \ldots, u_p^{(i)}) \in Q$, i=1,2 and $u_j^{(1)} \leq u_j^{(2)}$; j= 1,...,p, then $0 \leq \varphi(u_1^{(1)}, \ldots, u_p^{(1)}) \leq \varphi(u_1^{(2)}, \ldots, u_p^{(2)})$.

Let $\varphi \in \Gamma^3(Q)$, $Q = J_1 \times J_2 \times J_3$. Then the solution $\{t_k\}$ of the difference equation (2.4) is said to exist for given t_o, t_1 if

$$t_{k+1} - t_k \in J_1 \, , \quad t_k \in J_2 \cap J_3 \tag{2.5}$$

for all $k \geq 0$, i.e., if the entire sequence $\{t_k\}$ defined by (2.4) exists.

Using this notation we can formulate the following simple but, at the same time, general convergence theorem for the process (2.3).

<u>2.3</u> - <u>Basic Majorant Theorem</u>: Let $G: D \subset X \to X$ be an operator on the complete metric space X, and suppose there exists a function $\varphi \in \Gamma^3(Q)$ and a point $x_o \in D$ such that on some set $D_o \subset D$

$$(2.6) \qquad \rho(G(Gx),\ Gx) \leq \varphi(\ \rho(Gx,x),\ \rho(Gx,x_o),\ \rho(x,x_o))$$

whenever $x, Gx \in D_o$. Assume that for $t_o=0$, $t_1 = \rho(Gx_o,x_o)$ the solution $\{t_k\}$ of the difference equation (2.4) exists. If the sequence $\{x_k\}$ defined by (2.3) is contained in D_o, then $\{t_k\}$ majorizes $\{x_k\}$. Hence if $\lim t_k = t^* < +\infty$, also $\lim x_k = x^*$ exists, and the error estimate (2.2) holds. If $x^* \in D$ and G is continuous at x^*, then $x^* = Gx^*$.

The proof follows by induction. The relations

$$\rho(x_k,\ x_{k-1}) \leq t_k - t_{k-1}\ , \qquad \rho(x_k,\ x_o) \leq t_k$$

clearly hold for k=1, and for general k the isotonicity of φ implies that

$$\begin{aligned}\rho(x_{k+1},\ x_k) &\leq \varphi(\ \rho(x_k,\ x_{k-1}),\ \rho(x_k,\ x_o),\ \rho(x_{k-1},\ x_o)) \\ &\leq \varphi(t_k - t_{k-1},\ t_k,\ t_{k-1}) = t_{k+1} - t_k\end{aligned}$$

and hence by (2.1) that $\rho(x_{k+1},x_o) \leq t_{k+1} - t_o = t_{k+1}$. Thus, $\{x_k\}$ is majorized by $\{t_k\}$ which proves the convergence result. The fixpoint statement follows directly from (2.3) using the continuity of G at x^*.

In the application of this theorem it must be ascertained that $\{x_k\} \subset D_o$. A frequently used condition is $GD' \subset D'$ for some set $D' \subset D_o$; then clearly $\{x_k\} \subset D'$ whenever $x_o \in D'$. A conceptually different approach places restrictions on some of the early iterates and infers from this that all subsequent iterates remain in D_o.

<u>2.4</u> - <u>Lemma</u>: Let the conditions of Theorem <u>2.3</u> be satisfied. If $x_1,\dots,x_m \in D_o$ and

$$(2.7) \qquad \bar{S}(x_m,\ t_k-t_m) \subset D_o,\quad k=m,m+1,\dots\ ,$$ [1]

then $x_k \in D_o$ for $k \geq m$. If $\lim t_k = t^* < +\infty$ exists, (2.7) is satisfied if, either (a) $\bar{S}(x_m,\ t^* - t_m) \subset D_o$, or (b) $S(x_m,\ t^* - t_m) \subset D_o$ and $t_k < t^*$ for $k \geq 0$.

The proof proceeds again by induction. If $x_o,\dots,x_k \in D_o$ for some

---

[1]) $\bar{S}(x,r)$ and $S(x,r)$ denote the closed and open ball with center x and radius r.

$k \geq m$, then x_{k+1} is still defined and, as in the proof of Theorem 2.3, it follows that $\rho(x_{j+1}, x_j) \leq t_{j+1} - t_j$ for $j=0,1,\ldots,k$. Therefore by (2.1), $\rho(x_{k+1}, x_m) \leq t_{k+1} - t_m$, and hence $x_{k+1} \in \bar{S}(x_m, t_{k+1} - t_m) \subset D_o$. The second part of the statement is immediate.

In most applications the index m in (2.7) is set equal to 0 or 1, and - more specifically - the simple conditions $\bar{S}(x_o, t^*) \subset D_o$ or $\bar{S}(x_1, t^* - t_1) \subset D_o$ are used.

Theorem 2.3 together with Lemma 2.4 contain as special case the usual contraction theorem. In that instance we have $\varphi(u) = \alpha u$ with $\alpha < 1$ and (2.6) is replaced by the stronger condition

$$\rho(Gy, Gx) \leq \varphi(\rho(y,x)) = \alpha \rho(y,x) \quad , x,y \in D_o .$$

It then follows immediately that $t^* = \alpha t_1/(1-\alpha)$, and hence the condition $\bar{S}(x_1, t^* - t_1) \subset D_o$ is equivalent to the well-known condition $\bar{S}(x_1, r) \subset D_o$ for $r \geq \alpha \rho(x_1, x_o)/(1-\alpha)$.

Theorem 2.3 places the burden of the convergence proof on the analysis of the behavior of the difference equation (2.4). The question arises whether general conditions for φ can be found which assure the convergence of $\{t_k\}$ for certain t_1. In the special case of the equation

$$(2.8) \qquad t_{k+1} - t_k = \varphi(t_k - t_{k-1}) \ , \ t_o = 0, \ t_1 \text{ given,}$$

several such results can be derived. We shall not go into details here. The principal tool of the analysis is in this case the following well-known lemma which will also be needed several times later on (see e.g., [9]).

2.5 - <u>Kantorovich Lemma</u>: Let $\psi : [t_o, s_o] \subset R^1 \to R^1$ be continuous and isotone, and $\psi(t_o) \geq t_o$, $\psi(s_o) \leq s_o$. Then the sequences $t_{k+1} = \psi(t_k)$, $s_{k+1} = \psi(s_k)$, $k=o,1,\ldots$, satisfy

$$t_o \leq t_k \leq t_{k+1} \leq \lim t_k = t^* \leq s^* = \lim s_k \leq s_{k+1} \leq s_k \leq s_o$$

where t^* is the smallest and s^* the largest fixpoint of ψ in $[t_o, s_o]$.

In the case of the full difference equation (2.4) no general convergence conditions are known. The analysis of this equation is considerably simplified if there exists a function $\psi : J \subset R^1 \to R^1$ such that $J \subset J_3$, $\psi(J) \subset J_2$, and $\psi(J) - J \subset J_1$, and that

$$(2.9) \qquad \varphi(\psi(t) - t,\ \psi(t),\ t) = \psi(\psi(t)) - \psi(t), \quad t \in J.$$

It is then easily seen that $\{t_k\} \subset J$ is a solution of (2.4) with $t_o=0$, $t_1 = \psi(0)$ if and only if

$$(2.10) \qquad t_{k+1} = \psi(t_k), \quad t_o = 0,\ k=0,1,\ldots \quad .$$

Accordingly, (2.10) shall be called a 'first integral' of the difference equation (2.4). If $t^* \in J$ and ψ is continuous at t^*, (2.10) implies that $t^* = \psi(t^*)$, and if ψ is isotone, it follows from Lemma 2.5 that $\psi(t) > t$ for $0 < t < t^*$, and that t^* is the smallest non-zero fixpoint of ψ in J.

Theorem 2.3 represents a generalization of a similar contraction theorem of Collatz [6] who essentially assumes instead of (2.6) that

$$(2.11) \qquad \rho(Gy, Gx) \leq \psi(\rho(y,x) + \rho(x,x_o)) - \psi(\rho(x,x_o)), \quad x,y \in D_o$$

where $\varphi(u,v) = \psi(u+v) - \psi(v)$ is of class $\Gamma^2(Q)$. Using our terminology, Collatz then proceeds to show that

$$(2.12) \qquad t_{k+1} = \psi(t_k) + \gamma, \quad t_o = 0, \quad \gamma \geq \rho(x_1,x_o),\ k=0,1,\ldots$$

is a majorizing sequence of $\{x_k\}$. This is evidently equivalent to the assumption that (2.12) represents a first integral of the difference equation (2.4).

Instead of (2.11), Schröder [18] considers more general conditions, as, for example,

$$(2.13) \qquad \rho(Gy,Gx) \geq \varphi(\rho(y,x),\ \rho(y,x_o),\ \rho(x,x_o),\ \rho(Gy,x),\ \rho(Gx,y))$$

where $\varphi \in \Gamma^5(Q)$. But this generality is not used; in fact, the additional assumption is made that a function $\psi = \psi(u)$ exists such that

$$\psi(v) - \psi(u) \geq \varphi(v-u,\ v,\ u,\ \psi(v)-u,\ \psi(u)-v) .$$

This implies that (2.11) holds, and since in all subsequent considerations Schröder only uses the majorizing sequence (2.12), his results are no more general than those of Collatz.

In connection with the iterative processes (1.2) we shall see that condition (2.6) frequently arises in a very natural way while the corresponding more restrictive condition

$$(2.14) \qquad \rho(Gy, Gx) \leq \varphi(\rho(y, x),\ \rho(y,x_o),\ \rho(x,x_o))$$

only applies in a much smaller domain. However, the use of this condition

(2.14) together with the existence of a first integral (2.10) of (2.4) permits the derivation of some uniqueness results for the fixpoint x^*.

The following uniqueness theorem is a slight generalization of a result of Collatz [6] .

2.6 - First Uniqueness Theorem: Suppose that the assumptions of Theorem 2.3 are valid except that instead of (2.6) the condition (2.14) holds for all $x,y \in D_o$. Suppose that $\psi : J \subset R^1 \to R^1$ defines a first integral (2.10) of the difference equation (2.4), and that $\lim t_k = t^* = \psi(t^*) \in J$. Then there is no other fixpoint of G in the set $D_o \cap \bar{S}(x_o, t^*)$ except possibly x^*.

Proof: Suppose that $y^* = Gy^* \in D_o \cap \bar{S}(x_o, t^*)$, then $\rho(y^*,x_o) \le t^* = t^* - t_o$ and by induction we see that $\rho(y^*, x_k) \le t^* - t_k$ for all $k \ge 0$. In fact, since $\{t_k\}$ majorizes $\{x_k\}$ it follows that

$$\begin{aligned}\rho(y^*, x_{k+1}) = \rho(Gy^*, Gx_k) &\le \varphi(\rho(y^*, x_k), \rho(y^*, x_o), \rho(x_k, x_o)) \\ &\le \varphi(t^* - t_k, t^*, t_k) = \psi(t^*) - \psi(t_k) = t^* - t_{k+1} .\end{aligned}$$

This implies that $x^* = \lim x_k = y^*$.

Note that the existence of the first integral is not fully used and that the theorem remains valid if only the weaker condition

$$\varphi(t^* - t_k, t^*, t_k) = t^* - t_{k+1}$$

holds. For $\varphi(u) = \alpha u$, $0 < \alpha < 1$, the theorem evidently provides the usual uniqueness result connected with the contraction theorem.

In the case when φ depends only on the first two variables we get a somewhat larger uniqueness domain.

2.7 - Second Uniqueness Theorem: Assume that all assumptions of the first uniqueness theorem 2.7 hold except that the condition (2.14) is replaced by

$$(2.15) \qquad \rho(Gy, Gx) \le \varphi(\rho(y,x), \rho(y,x_o)) \quad , x,y \in D_o.$$

Suppose there exists a point $\hat{t} \in J$, $\hat{t} > t^*$ such that $\psi(t) < t$ for $t^* < t < \hat{t}$. Then there is no other fixpoint of G in $D_o \cap S(x_o, \hat{t})$ except possibly x^*.

Proof: Let $y^* = Gy^* \in D_o \cap S(x_o, \hat{t})$. If $s_o = \rho(y^*,x_o) \le t^*$, then the result is covered by 2.6 . Hence, assume that $t^* < s_o < \hat{t}$. It then follows by induction that $\rho(y^*, x_k) \le s_k - t_k$ where $s_{k+1} = \psi(s_k)$, $k=0,1,\dots$. In fact, using this as induction hypothesis we have

$$\rho(x_{k+1}, y^*) \leq \varphi(\rho(x_k, y^*), \rho(x_k, x_o)) \leq \varphi(s_k - t_k, t_k)$$
$$= \psi(s_k) - \psi(t_k) = s_{k+1} - t_{k+1} .$$

By Lemma 2.5, clearly $\lim s_k = t^*$ and hence again $x^* = \lim x_k = y^*$.

Note that the 'best possible' $\hat{t}$ is evidently the smallest fixpoint $t^{**} > t^*$, $t^{**} \in J$ of ψ , provided that such a fixpoint exists and of course that $\psi(t) < t$ for $t^* < t < t^{**}$.

3. A Basic Application

For the discussion in this and the following Sections we shall assume that, unless otherwise specified, X and Y denote real Banach spaces, and that $F: D \subset X \to Y$ is a given operator. Further, $L(X,Y)$ is the Banach space of all bounded linear operators with domain X and range in Y. For the sake of simplicity we shall also use the following notations:

(a) $F \in \mathrm{Lip}_\gamma(D_0)$, if $\|Fy - Fx\| \leq \gamma \|y - x\|$ for $x \in D_0 \subset D$.

(b) $F \in \mathcal{G}(D_0)$, if F possesses a bounded linear Gateaux derivative $\partial F(x) \in L(X,Y)$ for all $x \in D_0$.

(c) $F \in \mathcal{F}(D_0)$, if F has a Frechet derivative $F'(x)$ for all $x \in D_0$. Clearly then also $F \in \mathcal{G}(D_0)$.

Note that by definition $F \in \mathcal{G}(D_0)$ implies that every point $x \in D_0$ is an internal point of D, i.e., for $x \in D_0$ and $z \in X$ there exists an $\varepsilon > 0$ such that $x + tz \in D$ for $|t| \leq \varepsilon$.

The results collected in the following lemma are well-known; see, for example, Vainberg [19].

3.1 - If $F: D \subset X \to Y$ and $F \in \mathcal{G}(D_0)$ on some convex set $D_0 \in D$, then

(a) $\|\partial F(x)\| \leq \gamma$ for $x \in D_0$ implies that $F \in \mathrm{Lip}_\gamma(D_0)$;

(b) $\|\partial F(y) - \partial F(x)\| \leq \gamma$ for $x,y \in D_0$ implies that $\|Fy - Fx - \partial F(z)(y-x)\| \leq \gamma \|y - x\|$ for $x,y,z \in D_0$;

(c) $F \in \mathrm{Lip}_\gamma(D_0)$ implies that $F \in \mathcal{F}(D_0)$ and that $\|Fy - Fx - F'(x)(y-x)\| \leq \gamma \|y - x\|^2$ for $x,y \in D_0$.

A first and simple application of the results of Section 2 is the following theorem which will provide the basis for a number of subsequent uniqueness results.

3.2 - Theorem: Let $G: D \subset X \to X$ be such that $G \in \mathcal{F}(D_0)$ and $G' \in \mathrm{Lip}_\gamma(D_0)$ on some convex set $D_0 \subset D$. Assume that for some $x_0 \in D_0$ the estimates $\|G'(x_0)\| \leq \delta < 1$, $\|x_0 - Gx_0\| \leq \alpha$ and $h = \gamma\alpha/(1-\delta)^2 \leq 1/2$ hold, and set

$$(3.1) \qquad t^* = \frac{1 - \sqrt{1-2h}}{h} \, \frac{\alpha}{1-\delta}, \quad t^{**} = \frac{1 + \sqrt{1-2h}}{h} \, \frac{\alpha}{1-\delta}$$

Then, if $\bar{S}(x_0, t\ ) \subset D_0$, the iterates $x_{k+1} = Gx_k$, $k=0,1,\ldots$, remain in

$\bar{S}(x_o,t^*)$ and converge to a fixpoint x^* of G which is unique in $D_o \cap S(x_o,t^{**})$.

<u>Proof</u>: For $x,y \in D_o$ we have

$$\begin{aligned}\|Gy - Gx\| &\le \|Gy - Gx - G'(x)(y-x)\| + \|(G'(x)-G'(x_o))(y-x)\| \\ &+ \|G'(x_o)(y-x)\| \le \tfrac{1}{2}\gamma\|y-x\|^2 + \gamma\|x-x_o\|\|y-x\| \\ &+ \delta\|y-x\| = \varphi(\|y-x\|, \|x-x_o\|)\end{aligned} \tag{3.2}$$

where $\varphi(u,v) = \frac{1}{2}\gamma u^2 + \gamma vu + \delta u$. It is readily seen that $\varphi(u-v,v) = \psi(u) - \psi(v)$ where $\psi(t) = \frac{1}{2}\gamma t^2 + \delta t + \alpha$, and ψ is evidently isotone and has the fixpoints t^* and t^{**} . Moreover, $\psi(t) < t$ for $t^* < t < t^{**}$ unless $h = 1/2$ in which case $t^{**} = t^*$. Hence ψ defines a first integral of the difference equation (2.4) corresponding to φ, and the Kantorovich Lemma 2.5 assures that $\lim t_k = t^* < +\infty$. Now Lemma 2.4 implies that $x_k \in \bar{S}(x_o,t^*) \subset D_o$, and Theorem 2.3 provides the convergence statement. Since G is continuous, clearly $x^* = Gx^*$, and, in the case h=1/2, the uniqueness follows from Theorem 2.7 and otherwise from Theorem 2.8.

Note that in the case $h < 1/2$ also the usual contraction theorem applies. Thus the nonlinear estimate (3.2) provides here only the convergence for the border case $h = 1/2$ and gives the larger uniqueness domain.

This theorem has immediate application to the generalized chord method

$$x_{k+1} = x_k - A^{-1}Fx_k \quad , k=0,1,\ldots \tag{3.3}$$

where $A \in L(X,Y)$. For future reference we phrase this in form of the following corollary.

3.3 - For $F:D \subset X \to Y$ let $F \in \mathcal{T}(D_o)$ and $F' \in Lip_\gamma(D_o)$ on a convex set $D_o \subset D$. Suppose that $A \in L(X,Y)$ has a bounded inverse $A^{-1} \in L(Y,X)$ and $\|A^{-1}\| \le \beta$. Choose $x_o \in D_o$ such that $\|I - A^{-1}F'(x_o)\| \le \delta < 1$, $\|A^{-1}Fx_o\| \le \alpha$ and $h = \beta\gamma\alpha/(1-\delta)^2 \le 1/2$ and define t^*, t^{**} by (3.1). If $\bar{S}(x_o,t^*) \subset D_o$, then the iterates (3.3) remain in $\bar{S}(x_o,t^*)$ and converge to a solution x^* of $Fx = 0$ which is unique in $D_o \cap S(x_o,t^{**})$.

This result is essentially a theorem of Kantorovich and Akilov [9] . It contains as special case the well-known convergence theorem for the modified Newton method

$$x_{k+1} = x_k - (F'(x_o))^{-1}Fx_k \quad , k=0,1,\ldots \tag{3.4}$$

Another corollary is the following result of Linkov [12] .

3.4 - Let X be a real Hilbert space and $F: D \subset X \to X$ such that $F \in \mathfrak{F}(D_o)$ and $F' \in Lip_\gamma(D_o)$ on a convex set $D_o \subset D$. Suppose that for some $x_o \in D_o$, $F'(x_o)$ is selfadjoint and $\tau \|z\|^2 \geq (F'(x_o)z, z) \geq \sigma \|z\|^2$, $z \in X$, with $\sigma > 0$, and that $h = \gamma \|Fx_o\|/\sigma^2 \leq 1/2$. With $\alpha/(1-\delta) = \|Fx_o\|/\sigma$ define t^*, t^{**} by (3.1). If $\bar{S}(x_o, t^*) \subset D_o$, then the sequence

$$x_{k+1} = x_k - \frac{2}{\sigma+\tau} Fx_k \quad , k=0,1,..$$

remains in $\bar{S}(x_o, t^*)$ and converges to the only solution x^* of $Fx = 0$ in $D_o \cap S(x_o, t^*)$.

The proof follows from Theorem 3.2 if we set $Gx = x - \omega Fx$ with $\omega > 0$. Then $\|G'(x_o)\| \leq \max(|1-\omega\tau|, |1-\omega\sigma|) = \delta_\omega$ and δ_ω assumes its minimum for $\omega = 2/(\sigma+\tau)$.

The corollaries 3.3 and 3.4 were originally proved under the assumption that F is twice Frechet differentiable and $\|F''(x)\| \leq \gamma$ in D_o. If we reduce the conditions on F by assuming only that $F \in \mathcal{G}(D_o)$ and $\|\partial F(y) - \partial F(x)\| \leq \gamma$ in D_o, then the estimate (3.2) for $Gx = x - A^{-1}Fx$ reduces to a linear contraction estimate. Although this is an extremely simple result it has been repeatedly announced in various contexts, and we shall not go into details here. Slightly more interesting is the case when G is not a standard contraction but an 'iterated contraction' in the sense of Theorem 2.3.

3.5 - Let $F: D \subset X \to Y$ be such that $F \in \mathcal{G}(D_o)$ and $\|\partial F(y) - \partial F(x)\| \leq \gamma$ for all x,y from a convex set $D_o \in D$. Suppose that for some $x_o \in D_o$ and $B \in L(Y,X)$ we have $B\partial F(x_o)B = B$. If $\|B\| \leq \beta$ and $\beta\gamma < 1$ as well as $\bar{S}(x_o, t^*) \subset D_o$ where $t^* = \|BFx_o\|/(1-\beta\gamma)$, then the sequence $x_{k+1} = x_k - BFx_k$, k=0,1,.. , remains in $\bar{S}(x_o, t^*)$ and converges to a solution x^* of $BFx = 0$.

<u>Proof</u>: Set $Gx = x - BFx$, then

$$\begin{aligned} \|G(Gx) - Gx\| &= \|-BF(Gx)\| = \|-B \; F(x_o)BFx - BF(Gx) + BFx\| \\ &= \|B\,[\partial F(x_o)(Gx-x) - F(Gx) + Fx]\| \leq \beta\gamma \; \|Gx - x\| \end{aligned}$$

whenever $x, Gx \in D_o$. Hence, Theorem 2.3 applies with $\varphi(u) = \beta\gamma u$.

The condition $B\partial F(x_o)B = B$ plays a central role in the theory of generalized inverses of $\partial F(x_o)$. Theorem 3.5 represents a modified and somewhat improved version of a result of Ben-Israel [3] .

4. Approximate Newton Processes

In this Section the theory of Section 2 is applied to processes of the form (1.2). More specifically, we shall first consider the iterations

$$(4.1) \qquad x_{k+1} = x_k - A^{-1}(x_k)\, Fx_k \quad , \quad k=0,1,\ldots$$

where for fixed x, A(x) is a linear operator. Different results are obtained depending on the invertibility assumptions placed upon A(x); the simplest case is when A(x) has a bounded linear inverse in the entire domain.

4.1 - Theorem: Let $F: D \subset X \to Y$ be such that $F \in \mathfrak{J}(D_o)$ and $F' \in Lip_\gamma(D_o)$ on a convex set $D_o \subset D$. Suppose that $A: D_o \subset X \to L(X,Y)$ has for each $x \in D_o$ a bounded inverse $A^{-1} \in L(Y,X)$, and that $\|A^{-1}\| \leq \beta$, $\|F'(x) - A(x)\| \leq \delta$ for $x \in D_o$. Let $x_o \in D_o$ be such that $\|A^{-1}(x_o)Fx_o\| \leq \alpha$ and $h = \frac{1}{2}\beta\gamma\alpha + \beta\delta < 1$. If $\bar{S}(x_o,r) \in D_o$ where $r = \alpha/(1-h)$ then the sequence $\{x_k\}$ defined by (4.1) remains in $\bar{S}(x_o,r)$ and converges to a solution x^* of $Fx = 0$.

Proof: Define $G: D_o \subset X \to X$, $Gx = x - A^{-1}(x)Fx$; then, whenever $x, Gx \in D_o$

$$(4.2) \qquad \begin{aligned} \|G(Gx) - Gx\| &= \|-A^{-1}(Gx)F(Gx)\| \leq \beta\|F(Gx) - Fx - F'(x)(Gx-x)\| \\ &\quad + \beta\|(A(x)-F'(x))(Gx-x)\| \leq \varphi(\|Gx-x\|) \end{aligned}$$

where $\varphi(u) = \frac{1}{2}\beta\gamma u^2 + \beta\delta u$. Thus we have to consider the difference equation

$$(4.3) \qquad t_{k+1} - t_k = \frac{1}{2}\beta\gamma(t_k - t_{k-1})^2 + \beta\delta(t_k - t_{k-1}) \quad , \ k=0,1,\ldots$$

with $t_o = 0$, $t_1 = \alpha$. Evidently, $t_2 - t_1 = h\alpha < \alpha$ and by induction $t_{k+1} - t_k \leq h(t_k - t_{k-1})$ as well as $t_k - t_{k-1} \leq \alpha$ for all $k \geq 1$. Hence, $t_k \leq \alpha \sum_{j=0}^{k} h^j$ and therefore $\lim t_k = t^* \leq \alpha/(1-h) = r < +\infty$ exists and the convergence follows from Theorem 2.3. Since clearly $x^* \in D_o$ and $A^{-1}(x^*)$ is nonsingular, evidently $x^* = Gx^*$ implies $Fx^* = 0$.

In the special case of Newton's method we have $A(x) \equiv F'(x)$ and hence $\delta = 0$ and $h = \frac{1}{2}\beta\gamma$. For the solution of the reduced difference equation (4.3) it then follows immediately by induction that

$$t_{k+1} - t_k \leq \alpha\, h^{2^k - 1}$$

which readily implies that

$$t^* - t_k \leq \alpha \frac{h^{2^k-1}}{1 - h^{2^k}}$$

thus giving the usual quadratic convergence of Newton's method. This result about Newton's method is generally known as the Newton-Mysovskii theorem (see [13]).

In the general case when $\delta > 0$, the convergence is clearly only linear. Besides the error estimate (2.2) which always holds when a majorizing sequence has been found, we can also derive in exactly the same manner as (4.2) the non-computable error estimate

$$\| x^* - x_{k+1} \| \leq \frac{1}{2} \beta\gamma \| x^* - x_k \|^2 + \beta\delta \| x^* - x_k \|$$

which for $\delta = 0$ again gives the quadratic convergence of Newton's method.

It is not difficult to generalize this result - for example, by assuming only that F' satisfies a Hölder condition $\|F'(y) - F'(x)\| \leq \gamma \|y - x\|^{\lambda}$ with $0 < \lambda \leq 1$ for all $x, y \in D_o$. This changes the difference equation (4.3) to

$$t_{k+1} - t_k = \frac{1}{2} \beta\gamma (t_k - t_{k-1})^{1+\lambda} + \beta\delta (t_k - t_{k-1})$$

but does not affect the proof procedure. For $\lambda = 0$ we can use Lemma 3.1 (b) and obtain, already in the case when $F \in \mathcal{G}(D_o)$, that

$$\| G(Gx) - Gx \| \leq \beta(\gamma+\delta) \| Gx - x \|$$

which means that the convergence condition then reduces to $\beta(\gamma+\delta) < 1$.

In his convergence proof of Newton's method Kantorovich assumes only that $F'(x)$ is invertible at x_o, then using Banach's lemma to assure the invertibility at all further iterates. This idea can be extended to the general process (4.1). It leads to a nonlinear difference equation for which no explicit solution appears to be known except for special values of the parameters. Following is a summary of the basic material about this difference equation.

4.2 - Consider

$$(4.4) \quad \begin{aligned} & t_{k+1} - t_k = \frac{1}{1 - p_4 t_k} \left[p_1 (t_k - t_{k-1})^2 + (p_2 + p_3 t_{k-1})(t_k - t_{k-1}) \right] \\ & t_o = 0 \ , \ t_1 = \alpha \ , \ 0 \leq p_4 \alpha < 1 \ , \qquad k=1,2,\ldots \end{aligned}$$

where $p_i \geq 0$, $i=1,\ldots,4$. If for some parameter set (p_i^o, α^o) the solution $\{t_k^o\}$ satisfies $t_k^o \leq t_{k+1}^o$, $k \geq 0$, and $\lim t_k = \hat{t}^* < 1/p_4$, then for any (p_i, α) with $0 \leq p_i \leq p_i^o$, $i=1,\ldots,4$, $0 \leq \alpha \leq \alpha^o$, the solution $\{t_k\}$ is again nondecreasing and $\lim t_k = t^* \leq \hat{t}^*$. If $p_1 > 0$, $0 \leq p_2 < 1$, $p_3 + p_4 = 2p_1$ and $0 < \alpha \leq (1-p_2)^2/4p_1$, then the t_k are strictly increasing and

$$\lim t_k = t^* = \frac{1}{2p_1}\left[(1 - p_2) - \sqrt{(1-p_2)^2 - 4p_1\alpha}\right]$$

Proof: Clearly,

$$\varphi(u,v,w) = \frac{1}{1 - p_4 v}\left[p_1 u^2 + (p_2 + p_3 w)\, u\right]$$

is of class $\Gamma^3(Q)$ with $Q = [0,\infty) \times [0, 1/p_4) \times [0,\infty)$. The fact that $t_1^o \geq t_{1_o} \geq 0$ and $t_k^o \leq t_{k+1}^o \leq \overset{o}{t}{}^* < 1/p_4$ then implies by induction that $t_k \leq t_k^o$ for $k \geq 0$.

For the proof of the second part set

$$u(t) = p_1 t^2 - (1 - p_2)t + \alpha \quad , \; v(t) = 1 - p_4 t \quad .$$

Then

$$t_{k+1} = t_k + \frac{u(t_k)}{v(t_k)} \quad , \; t_o = 0 \; , \; k=0,1,\ldots$$

is a first integral of (4.4). In fact,

$$(4.5) \qquad \begin{aligned} t_{k+1} - t_k &= \frac{1}{v(t_k)}\Big[u(t_k) - u(t_{k-1}) - u'(t_{k-1})(t_k - t_{k-1}) \\ &\quad + (u'(t_{k-1}) + v(t_{k-1}))(t_k - t_{k-1})\Big] = \varphi(t_k - t_{k-1}, t_k, t_{k-1}). \end{aligned}$$

It is easy to verify that t^* is the smallest fixpoint of $\psi(t) = t + u(t)/v(t)$. Moreover, since $u(t)/v(t) > 0$ for $0 < t < t^*$, it follows that $t < \psi(t)$ for $0 < t < t^*$. A computation analogous to (4.5) shows that

$$t^* - \psi(t) = \frac{1}{v(t)}\left[p_1 (t^* - t)^2 + (p_2 + p_3 t)(t^* - t)\right] > 0$$

for $0 < t < t^*$ since

$$p_4 t < p_4 t^* = (1-p_2) - \sqrt{(1-p_2)^2 - 4p_1\alpha} - p_3 t^* \leq 1.$$

Now $t_1 = \alpha < t^*$, for otherwise $u(\alpha) = 0$ leads to a contradiction, and hence $t_k < t_{k+1} < t^*$ for all $k \geq 0$ and $\lim t_k = t^*$.

Using this lemma we now obtain the following convergence result for (4.1).

4.3 - Theorem: For $F: D \subset X \to Y$ let $F \in \mathcal{T}(D_o)$ and $F' \in Lip_\gamma(D_o)$ in a convex set $D_o \subset D$. Let $A: D_o \subset X \to L(X,Y)$ and a point $x_o \in D_o$ be such that $\| A(x) - A(x_o) \| \leq \eta \| x - x_o \|$, and $\| F'(x) - A(x) \| \leq \delta_o + \delta_1 \| x - x_o \|$ for $x \in D_o$, $(\delta_o, \delta_1 \geq 0)$. Assume that $A(x_o)$ has a bounded inverse $A^{-1}(x_o) \in L(Y,X)$ with $\| A^{-1}(x_o) \| \leq \beta$, $\| A^{-1}(x_o) F x_o \| \leq \alpha$, and that $\beta\delta_o < 1$ and

$h = \sigma\beta\gamma\alpha/(1-\beta\delta_o)^2 \leq 1/2$ where $\sigma = \max(1, (\eta+\delta_1)/\gamma)$. Set

$$(4.6) \qquad t^* = \frac{1 - \sqrt{1-2h}}{h} \frac{\alpha}{1-\beta\delta_o}, \quad t^{**} = \frac{1 + \sqrt{1-2h/\sigma}}{h} \frac{\sigma\alpha}{1-\beta\delta_o} .$$

If $\bar{S}(x_o,t^*) \subset D_o$, then the sequence $\{x_k\}$ defined by (4.1) remains in $S(x_o,t^*)$ and converges to a solution x^* of $Fx = 0$ which is unique in $D_o \cap S(x_o,t^{**})$.

Proof: For $x \in S(x_o,t^*)$ we have

$$\| A(x) - A(x_o)\| \leq \eta \| x - x_o\| < \eta\, t^* \leq \sigma\gamma\, t \leq \frac{1 - \beta\delta_o}{\beta} \leq \frac{1}{\beta}$$

Hence, by Banach's lemma, $A(x)$ is nonsingular and

$$\| A^{-1}(x)\| \leq \frac{\beta}{1 - \beta\eta \|x - x_o\|}, \quad x \in S(x_o, t^*) .$$

Therefore, $Gx = x - A^{-1}(x)Fx$ is defined on $S(x_o,t^*)$, and if x, Gx are contained in this open ball, then

$$\| G(Gx) - Gx\| = \| - A^{-1}(Gx)F(Gx)\|$$

$$(4.7) \quad \leq \frac{\beta}{1 - \beta\eta \| Gx-x_o\|}\Big[\|F(Gx)-Fx-F'(x)(Gx-x)\| + \|(F'(x)-A(x))(Gx-x)\|\Big]$$

$$\leq \varphi(\| Gx-x\| , \| Gx-x_o\| , \| x-x_o\|)$$

where

$$\varphi(u,v,w) = \frac{1}{1 - \beta\eta v}\left[\frac{1}{2}\beta\gamma u^2 + \beta(\delta_o + \delta_1 w)\, u\right]$$

Hence, the difference equation (2.4) in this case has the form (4.4). Now,

$$0 < p_1 = \frac{1}{2}\beta\gamma \leq p_1^o = \frac{1}{2}\sigma\beta\gamma , \quad p_2 = p_2^o = \beta\delta_o < 1 ,$$

$$p_3 = \beta\delta_1 \leq p_3^o = \beta(\sigma\gamma - \eta) ; \quad p_4 = p_4^o = \beta\eta$$

and

$$\alpha = \alpha^o \leq \frac{1}{2p_1}(1 - p_2^o)^2 .$$

The case $\alpha = 0$ can be excluded since otherwise already $Fx_o = 0$; by 4.2 it therefore follows that the t_k are strictly increasing and that $\lim t_k \leq t^*$. Now by Lemma 2.4 (b), $\{x_k\} \subset S(x_o,t^*)$ and convergence is obtained by Theorem 2.3. The fact that $Fx^* = 0$ follows from

$$(4.8) \quad \begin{aligned} \| Fx_k\| &\leq \|(A(x_k)-A(x_o))(x_{k+1}-x_k)\| + \| A(x_o)(x_{k+1}-x_k)\| \\ &\leq (\eta\, t^* + \|A(x_o)\|)\, \|x_{k+1} - x_k\| \end{aligned}$$

The uniqueness is a direct consequence of 3.3 applied to the modified process $x_{k+1} = x_k - A^{-1}(x_o)Fx_k$, $k=0,1,\ldots$.

The frequently mentioned Newton-Kantorovich theorem is obtained as the special case when $F'(x) \equiv A(x)$. Then $\gamma = \eta$, $\delta_o = \delta_1 = 0$ and hence $\sigma = 1$, and all conditions reduce exactly to those of the Newton-Kantorovich theorem. A direct proof of that theorem using the estimate (4.7) and the concept of majorizing sequences was independently found by Ortega [16], showing the simplicity of this type of convergence proof when compared with the proofs known heretofore.

Note, if we assume only that $\|F'(x_o) - A(x_o)\| \leq \delta_o$, then the other conditions on F' and A assure that $\|F'(x) - A(x)\| \leq \delta_o + (\gamma+\eta)\|x - x_o\|$. In that case we need $\sigma = 3$.

As always, the error estimate (2.2) is available for the iteration considered in 4.3. In addition, a non-computable error estimate of the form

$$\|x^* - x_{k+1}\| \leq \frac{\beta}{1-\beta\eta\|x_k-x_o\|}\left[\frac{1}{2}\gamma\|x^*-x_k\|^2 + (\delta_o+\delta_1\|x_{k-1}-x_o\|)\|x^*-x_k\|\right]$$

can be derived in the same way as the principal estimate (4.7). For Newton's method this provides in the case $h < 1/2$, i.e., $t^* < 1/\beta\eta$, the estimate

$$\|x^* - x_{k+1}\| \leq \frac{1}{2}\frac{\beta\gamma}{1-\beta\eta t^*}\|x^* - x_k\|^2,$$

again giving the quadratic convergence. If δ_o, δ_1 are not both zero, then $t^* < 1/\beta\eta$ and hence we have linear convergence with a convergence factor $(\beta\delta_o+\beta\delta_1 t^*)/(1-\beta\eta t^*)$.

Another special case of Theorem 4.3 is the following extension of a theorem of Bryan [4].

4.4 - For $F: D \subset X \to Y$ let $F \in (D_o)$ and $F' \in Lip_\gamma(D_o)$ on some convex set $D_o \subset D$, and suppose that $P: L(X,Y) \to L(X,Y)$ is a bounded linear operator with $\|P\| \leq 1$ and $\|I-P\| \leq 1$. Assume that at $x_o \in D_o$, $(PF'(x_o))^{-1} \in L(Y,X)$ exists and that the estimates $\|(PF'(x_o))^{-1}\| \leq \beta$, $\|F'(x_o)\| \leq \delta_o$ $\|(PF'(x_o))^{-1}Fx_o\| \leq \alpha$, and $\beta\delta_o < 1$, $h = 2\beta\gamma\alpha/(1-\beta\delta_o)^2 \leq 1/2$ hold. With $\sigma = 2$ define t^*, t^{**} by (4.6). If $\bar{S}(x_o,t^*) \subset D_o$, then the sequence

$$x_{k+1} = x_k - (PF'(x_k))^{-1}Fx_k, \quad k=0,1,\ldots$$

remains in $S(x_o,t^*)$ and converges to a solution x^* of $Fx = 0$ which is unique in $D_o \cap S(x_o,t^{**})$.

The proof follows immediately from Theorem 4.3 if we set $A(x) \equiv PF'(x)$; then $\gamma = \eta$, $\delta_1 = \gamma$, and $\sigma = 2$.

Bryan developed his result for a convergence analysis of the Newton-Jacobi iteration

$$x_1^{(k+1)} = x_i^{(k)} - \frac{f_i(x^{(k)})}{\partial^i f_i(x^{(k)})} \quad , \ i=1,\dots,n \ , \ k=0,1,\dots$$

for the solution of a nonlinear system $f_i(x_1,\dots,x_n) = 0$, $i=1,\dots,n$; this method was originally suggested by Lieberstein [11]. In this case P maps every $n \times n$ matrix $A = (a_{ij})$ into its diagonal part $PA = \mathrm{diag}(a_{11},\dots,a_{nn})$, and monotone norms have to be used.

Zincenko [21], [22] has shown that the differentiability condition on F used in Theorems 4.1 and 4.3 can be replaced by corresponding conditions on A. His theorems, originally proved by Kantorovich's majorant method, can also be proved easily by the techniques developed here. We shall phrase only the Zincenko result corresponding to Theorem 4.3 ; the result corresponding to the simpler Theorem 4.1 should then be self-evident.

4.5 - Suppose that on some convex set $D_o \subset D_F \cap D_K$, $F: D_F \subset X \to Y$ is continuous and $K: D_K \subset X \to Y$ satisfies $K \in \mathfrak{F}(D_o)$, $K' \in \mathrm{Lip}_\gamma(D_o)$, and, moreover, that $F-K \in \mathrm{Lip}_\delta(D_o)$. Assume that for some $x_o \in D_o$, $(K'(x_o))^{-1} \in L(Y,X)$ exists and that $\|(K'(x_o))^{-1}\| \le \beta$, $\|(K'(x_o))^{-1}Fx_o\| \le \alpha$, as well as $\beta\delta < 1$, and $h = \beta\gamma\alpha/(1-\beta\delta)^2 \le 1/2$. With $\delta_o = \delta$ and $\sigma = 1$ define t^*, t^{**} by (4.6). If $\bar{S}(x_o,t^*) \subset D_o$, then the sequence

$$x_{k+1} = x_k - (K'(x_k))^{-1}Fx_k \quad , \ k=0,1,\dots$$

remains in $S(x_o,t^*)$ and converges to the only solution x^* of $Fx = 0$ in $D_o \cap S(x_o,t^{**})$.

Proof: For $x \in S(x_o,t^*)$ we have

$$\| K'(x) - K'(x_o)\| \le \gamma \|x - x_o\| < \gamma t^* \le \frac{1}{\beta}$$

and hence by Banach's lemma, $K'(x)$ is nonsingular and

$$\|(K'(x))^{-1}\| \leq \frac{\beta}{1-\beta\gamma\|x-x_o\|} \quad , \; x \in S(x_o,t^*) \; .$$

Therefore, $Gx = x - (K'(x))^{-1}Fx$ is defined on the open ball $S(x_o,t^*)$ and, if x, Gx are in this ball, then

$$\| G(Gx) - Gx \| = \| -(K'(Gx))^{-1}F(Gx) \|$$

$$\leq \frac{\beta}{1-\beta\gamma \|Gx-x_o\|}\Big[\|K(Gx)-Kx-K'(x)(Gx-x)\| + \|(F(Gx)-K(Gx)) - (Fx-Kx)\| \Big]$$

$$\leq \varphi(\| Gx-x \| , \| Gx-x_o\|)$$

where

$$\varphi(u,v) = \frac{\beta}{1-\beta\gamma v}\Big[\frac{1}{2}\gamma u^2 + \delta u\Big].$$

The difference equation is therefore a special case of (4.4) with

$$p_1 = \frac{1}{2}\beta\gamma \;, \; p_2 = \beta\delta \;, \; p_3 = 0, \;\; p_4 = \beta\gamma = 2p_1 \;.$$

This falls under the case considered in the second part of 4.2 and the convergence statement in the theorem is now a direct consequence of Theorem 2.3 together with Lemma 2.4 (b). The fact that $Fx^* = 0$ follows from the estimate formed analogously to (4.8); finally, the uniqueness is a direct consequence of 3.3 applied to the process $x_{k+1} = x_k - (K'(x_o))^{-1}Fx_k$, $k=0,1,\ldots$.

So far, we have considered only the iteration (4.1) and not the process (1.2) mentioned in the introduction. In the case when in (1.2) $B(x)$ no longer has a bounded linear inverse, most results given in the literature simply reduce to Theorem 3.1 . Of a slightly different character is the following generalization of a result of Ben-Israel [3].

4.6 - Let $F: D \subset X \to Y$ be such that $F \in \mathcal{F}(D_o)$ and $F' \in Lip_\gamma(D_o)$ on some convex set $D_o \subset D$, and suppose that $B: D_o \subset X \to L(Y,X)$ is a mapping with the properties $\|B(x)\| \leq \beta$, and $\|(B(y)-B(x))Fy\| \leq \eta \|y-x\|$ for $x,y \in D_o$. Moreover, let

(4.8) $\qquad \| B(x)F'(x)z - z\| \leq \delta \|z\|$ for $x \in D_o$, $z \in B(x)Y$.

If for some $x_o \in D_o$, $\|B(x_o)Fx_o\| \leq \alpha$, $h = \frac{1}{2}\beta\gamma\alpha + (\eta+\delta) < 1$ as well as $\bar{S}(x_o, \alpha/(1-h)) \subset D_o$, then the sequence $\{x_k\}$ defined by (1.2) remains in this ball and converges to a solution x^* of $B(x)Fx = 0$.

Proof: Set $G: D_o \subset X \to X$, $Gx = x - B(x)Fx$; then, for x, $Gx \in \bar{S}(x_o,\alpha/(1-h))$

$$\|G(Gx)-Gx\| \leq \|(B(x)-B(Gx))F(Gx)\| + \|B(x)(F(Gx)-Fx-F'(x))(Gx-x)\|$$

$$+ \|B(x)F'(x)(Gx-x) - (Gx-x)\| \leq \frac{1}{2}\beta\gamma \|Gx-x\|^2 + (\eta+\delta)\|Gx - x\|$$

where it was taken into account that by definition $Gx-x \in B(x)Y$. Hence, we have a nonlinear estimate with $\varphi(u) = \frac{1}{2}\beta\gamma u^2 + (\eta+\delta)u$. The corresponding difference equation has already been considered in the proof of Theorem 4.1 and it is now easy to check that the result follows directly from Theorem 2.3.

Note that when condition (4.8) is replaced by the stronger condition $\| I - B(x)F'(x)\| \leq \delta$, $x \in D_o$, we obtain $\|Gy - Gx\| \leq \varphi(\|y - x\|)$. On the other hand, if we weaken the differentiability condition for F by assuming only that $F \in \mathcal{G}(D_o)$ and $\|\partial F(y) - \partial F(x)\| \leq \gamma$, then $\|G(Gx) - Gx\| \leq (\beta\gamma+\eta+\delta)\|Gx - x\|$ for $x, Gx \in D_o$. This represents the approach used by Ben-Israel.

5. Some Generalizations

In its present form the theory of Section 2 applies only to processes of the form $x_{k+1} = Gx_k$, $k=0,1,\ldots$. It is not difficult to see how this theory can be extended to cover the more general processes

$$(5.1) \qquad x_{k+1} = G_k x_k \quad , \; k=0,1,\ldots$$

which include for instance the iterations (1.3).

5.1 - Extended Majorant Theorem: Consider a sequence of operators $G_k : D_k \subset X \to X$, $k=0,1,\ldots$, on the complete metric space X, and suppose that with certain $\varphi_k \in \Gamma^3(Q)$, $k=0,1,\ldots$, and some point $x_o \in D_o$ the estimates

$$(5.2) \qquad \rho(G_{k+1}(G_k x), G_k x) \le \varphi_k(\rho(G_k x, x), \rho(G_k x, x_o), \rho(x, x_o)) \; , \; k=0,1,\ldots$$

hold whenever $x, G_k x$ belong to a convex set $D \subset \bigcap_{k=0}^{\infty} D_k$. Assume further that the sequence $\{t_k\}$ defined by the recursion

$$(5.3) \qquad t_{k+1} - t_k = \varphi_{k-1}(t_k - t_{k-1}, t_k, t_{k-1}), \; k=1,2,\ldots,$$

with $t_o = 0$ and $t_1 = \rho(G_o x_o, x_o)$ exists and converges to $t^* < +\infty$. If the sequence $\{x_k\}$ defined by (5.1) remains in D, then $\{t_k\}$ majorizes $\{x_k\}$ and hence $\lim x_k = x^*$ exists and the error estimate (2.2) holds.

The proof is completely analogous to that of Theorem 2.3.

The condition $\{x_k\} \subset D$ can again be replaced by the assumptions of Lemma 2.4, or by a condition such as $G_k D' \subset D'$ for all $k \ge 0$ with $D' \subset D$.

If (5.2) is changed to the stronger estimate

$$(5.4) \qquad \rho(G_{k+1} y, G_k x) \le \varphi_k(\rho(y,x), \rho(y,x_o), \rho(x,x_o)) \quad , \; k=0,1,\ldots$$

then a fixpoint statement can be obtained.

5.2 - Suppose all conditions of Theorem 5.1 are valid but that instead of (5.2) the condition (5.4) holds for all $x,y \in D$. Moreover, let

$$(5.5) \qquad \lim_{k\to\infty} \varphi_k(t^* - t_k, t^*, t_k) = 0 \; .$$

If $\hat{G}: \hat{D} \subset X \to X$ is an operator such that $x^* \in \hat{D}$ and that

$$\lim_{k\to\infty} \rho(\hat{G}x^*, G_k x^*) = 0 \quad ,$$

then x^* is a fixpoint of $\hat{G}$.

The proof follows directly from

$$\begin{aligned}\rho(\hat{G}x^*, x^*) &\leq \rho(\hat{G}x^*, G_{k+1}x^*) + \rho(G_{k+1}x^*, G_k x_k) + \rho(x_{k+1}, x^*) \\ &\leq \rho(\hat{G}x^*, G_{k+1}x^*) + \varphi_k(\rho(x^*,x_k), \rho(x^*,x_o), \rho(x_k,x_o)) + \rho(x_{k+1}, x^*) \\ &\leq \rho(\hat{G}x^*, G_{k+1}x^*) + \varphi_k(t^*-t_k, t^*,t_k) + \rho(x_{k+1},x^*) .\end{aligned}$$

Note that in the case $\varphi_k \equiv \varphi$, $k \geq 0$, condition (5.5) will be satisfied if φ is continuous at $(0,t^*,t^*)$.

5.2 complements results for iterations of the type (5.1) obtained by Ortega and Rheinboldt [14].

Similar to the results of Sections 3 and 4 we can now formulate a variety of results for iterations of the type

$$(5.6) \qquad x_{k+1} = x_k - A_k^{-1}Fx_k \ , \ k=0,1,\ldots$$

As a typical example, an extension of Theorem 4.1 shall be presented here. X and Y now again denote Banach spaces.

5.3 - For $F: D \subset X \to Y$ let $F \in \mathcal{J}(D_o)$ and $F' \in Lip_\gamma(D_o)$ on a convex set $D_o \subset D$, and suppose that $A_k \in L(X,Y)$, $k=0,1,\ldots$, is a sequence of mappings with bounded inverses $A_k^{-1} \in L(Y,X)$ and $\|A_k^{-1}\| \leq \beta_k \leq \beta$. Let $\|F'(x_o)-A_k^{-1}\| \leq \delta_k$ and $\beta_{k+1}\delta_k \leq \delta < 1$, and assume that for some $x_o \in D_o$, $\|A_o^{-1}Fx_o\| \leq \alpha$ and $h = \beta\gamma\alpha/(1-\delta)^2 \leq 1/2$. Define t^* and t^{**} by (3.1). If $\bar{S}(x_o,t^*) \subset D_o$ then the sequence $\{x_k\}$ defined by (5.6) remains in $\bar{S}(x_o,t^*)$ and converges to a solution x^* of $Fx = 0$ which is unique in $D_o \cap S(x_o,t^{**})$.

For the proof set $G_k x = x - A_k^{-1}Fx$; then for $x, G_k x \in D_o$

$$\begin{aligned}\|G_{k+1}(G_k x)- G_k x\| &= \|-A_{k+1}^{-1}(F(G_k x)\| \leq \beta_{k+1}\|F(G_k x)-Fx-F'(x)(G_k x-x)\| \\ &\quad + \beta_{k+1}\left[\|F'(x)-F'(x_o)\| + \|F'(x_o)-A_k^{-1}\|\right]\|G_k x - x\| \\ &\leq \frac{1}{2}\beta\gamma\|G_k x - x\|^2 + (\delta + \beta\gamma\|x - x_o\|)\|G_k x - x\|\end{aligned}$$

which leads to a difference equation of the same form as that used in the proof of Theorem 3.1. Thus the result is a direct consequence of Lemma 2.4 and Theorem 5.1. The uniqueness follows again from Theorem 3.1 applied to the process $x_{k+1} = x_k - A_o^{-1}Fx_k$, $k=0,1,\ldots$.

If we reduce the differentiability condition of F to $F \in \mathcal{G}(D_o)$ and assume only that $\| \partial F(y) - \partial F(x) \| \leq \gamma$ in D_o, then (5.7) reduces to

$$(5.8) \qquad \| G_{k+1}(G_k x) - G_k x \| \leq (\beta\gamma + \delta) \| G_k x - x \| .$$

The corresponding convergence theorem represents essentially a result of Bartle [1].

It is even possible to drop the differentiability of F entirely by assuming only that

$$(5.9) \qquad \| Fy - Fx - A_k(y-x) \| \leq \eta_k \| y - x \| , \quad x,y \in D_o , \quad k=0,1,..$$

where $\beta_{k+1}\eta_k \leq h < 1$ and $\bar{S}(x_o, \alpha/(1-h)) \in D_o$ and $\{\|A_k\|\}$ is bounded.

Theorem 5.3 as well as the simpler results based on (5.8) or (5.9) can be applied, for instance, to approximate Newton processes of the form

$$x_{k+1} = x_k - (F'(z_k))^{-1} Fx_k , \quad k=0,1,..$$

where $\{z_k\}$ is some given sequence of points. Such processes have been considered by Bartle [1] and Schröder [18]. The various possible results depend in this case again on (a) the differentiability assumption about F, i.e., whether $F \in \mathcal{J}(D_o)$ and $F' \in Lip_\gamma(D_o)$ or $F \in \mathcal{G}(D_o)$ and $\|\partial F(y) - \partial F(x)\| \leq \delta$, and on (b) the invertibility assumption about $F'(x)$, i.e., whether $(F'(x))^{-1} \in L(Y,X)$ exists for all $x \in D_o$ or only $(F'(x_o))^{-1} \in L(Y,X)$. We shall not formulate here the different theorems arising from the various combinations of these conditons; their form as well as their proofs based on Theorem 5.1 should be evident.

There is also a possibility of applying Theorem 5.1 directly to processes of the form

$$x_{k+1} = x_k - A^{-1}(z_k) Fx_k , \quad k=0,1,..$$

When the same assumptions are made about A(x) as in Section 4, the resulting difference equations are again either of the form used in Theorem 4.1 or of that discussed in Lemma 4.2.

In Theorem 5.3 and the subsequent discussions we have avoided the difficult problem of analysing the recursion (5.3) by forcing all φ_k to be equal. The problem of determining when the solution of the recursion relation (5.3) with variable φ_k converges to a finite limit, is for all practical purposes an open question, except when (5.2) has the special form

$$(5.10) \qquad \rho(G_{k+1}(G_k x),\ G_k x) \ \le\ \eta_k\ \ \rho(G_k x,\ x)\ ,\ k=0,1,\ldots\ ,$$

i.e., when (5.3) reduces to

$$t_{k+1} - t_k = \eta_{k-1}(t_k - t_{k-1}),\ k=1,2,\ldots\ ,\ t_o = 0,\ t_1 = \alpha\ .$$

In that case we have

$$t_{k+1} - t_k = (\prod_{j=0}^{k} \eta_j)\ \alpha$$

and it is readily seen that $t^* < +\infty$ if $\eta_k \le \eta < 1$ for $k \ge k_o$. This can be applied to (5.8) when the β_k are not estimated by β, i.e., when the constant on the right is allowed to depend on k.

It should also be noted that the results of Kivistik [10] are of the type considered in Theorem 5.1 and that in one of his cases the φ_k are variable and are themselves given by a recursion relation.

Theorem 5.1 represents only one possible extension of Theorem 2.3 to the iterations (5.1). In particular, there is no reason why the right hand side of (5.2) should only depend on three terms and should not include terms such as $\rho(G_j x,\ G_{j-1}x)$ etc. with $1 \le j \le k-1$. This of course increases the order of the difference equation and thus makes it even more difficult to find convergence conditions for the t_k.

A very simple example of this extended type of estimate is a wellknown result of Cacciopoli [5] , and later Weissinger [20] , who considered the iteration $x_{k+1} = Gx_k$, k=0,1,.. , under the generalized contraction condition

$$\rho(G^k y,\ G^k x) \ \le\ \alpha_k\ \rho(y,\ x)$$

with $\sum_{k=0}^{\infty} \alpha_k < \infty$. In our setting this condition can be weakened to

$$\rho(G^{k+1}x,\ G^k x) \ \le\ \alpha_k\ \rho(Gx,\ x)\ ,\ \alpha_o = 1\ .$$

The corresponding difference equation is then

$$t_{k+1} - t_k = \alpha_k\, t_1\ ,\ t_o = 0\ ,\ t_1 = \alpha$$

with $t^* = (\sum_{k=0}^{\infty} \alpha_k)\alpha$.

Finally, it should be noted that all our results can be extended immediately to spaces metricized by elements of a partially ordered topological linear space. See, for example, Collatz [6] for a discussion of such spaces. However, in that case the difference equation (2.4) and the recursion (5.3) represent relations in such partially ordered spaces and the problem of determining when the t_k converge is compounded even further. A variety of special results can of course be formulated also in this case, but deeper results can only be expected once the resulting difference equations and recursions are better understood.

6. References

[1] Bartle, R.G., Newton's method in Banach spaces; Proc.A.M.S.,6,1955,827

[2] Ben-Israel,A., A modified Newton-Raphson method for the solution of systems of equations; Israel J.Math.,3,1965,94

[3] Ben-Israel,A., On the Newton-Raphson method;Northwestern Univ.,Syst. Res. Mem. 162, 1966

[4] Bryan,C.A., An iterative method for solving nonlinear systems of equations; Ph.D. Dissertation,Univ. of Arizona,1963

[5] Cacciopoli,R., Un teorema generale sull' existenca di elementi uniti in una traetor-maxione funzionale; Atti Accad.Naz.Lincei Rend. Ci. Csi.Fis.Mat.Nat., 6,11,1930,794

[6] Collatz,L., Funktionalanalysis und Numerische Mathematik; Springer, Heidelberg/Germany, 1965

[7] Kantorovich,L.V., Functional analysis and applied mathematics; Usp. Mat.Nauk,3,6,1948,89; transl from the Russian by C.D.Benster, U.S.Nat'l Bureau of Standards, 1953

[8] Kantorovich,L.V., On Newton's method; Trudy Mat. Inst.Steklov, 28, 1949,104

[9] Kantorovich,L.V. and Akilov,G.P., Functional analysis in normed spaces; transl. from the Russian, Pergamon Press, 1964

[10] Kivistik,L., A modification of the minimal-residual iteration method for the solution of equations involving nonlinear operators; Dokl.Akad.Nauk, 136, 1961, 22

[11] Lieberstein,H.M., Overrelaxation for nonlinear elliptic partial differential equations; M.R.C. Tech.Summary Report 80, 1959

[12] Linkov, E.I., The convergence of some iterative methods; Moscow Oblast.Ped. Inst. Ucen.Zap., 150, 1964, 71

[13] Mysovskii, I., On the convergence of the method of L.V.Kantorovich for the solution of functional equations and its applications; Dokl. Akad. Nauk, 70, 4, 1950, 565

[14] Ortega,J. and Rheinboldt,W., On a class of approximate iterative processes; Arch.f.Rat.Mech.and Analysis, 23,5,1967,352

[15] Ortega,J. and Rheinboldt,W., Monotone iterations for nonlinear equations with applications to Gauss-Seidel methods; SIAM J. on Num. Anal., 1967, to appear shortly

[16] Ortega,J., The Newton-Kantorovich theorem; submitted to Am.Math.Monthly

[17] Schröder,J., Nichtlineare Majoranten beim Verfahren der schrittweisen Näherung; Arch. Math. 12, 1956, 471

[18] Schröder,J., Über das Newtonsche Verfahren; Arch. Rat. Mech. Anal.,1, 1957, 154

[19] Vainberg,M., Variational methods in the study of nonlinear operators; transl. from the Russian by A.Feinstein, Holden-Day, 1964

[20] Weissinger,J., Zur Theorie und Anwendung des Iterationsverfahrens; Math. Nachr., 8, 1952, 193

[21] Zincenko,O.I., Some approximate methods of solving equations with non-differentiable operators; Dopovidi Akad.Nauk Ukrain RSR, 1963, 156

[22] Zincenko,O.I., A class of approximate methods of solving operator equations with non-differentiable operators; Dopovidi Akad. Nauk Ukrain RSR, 1963,852